

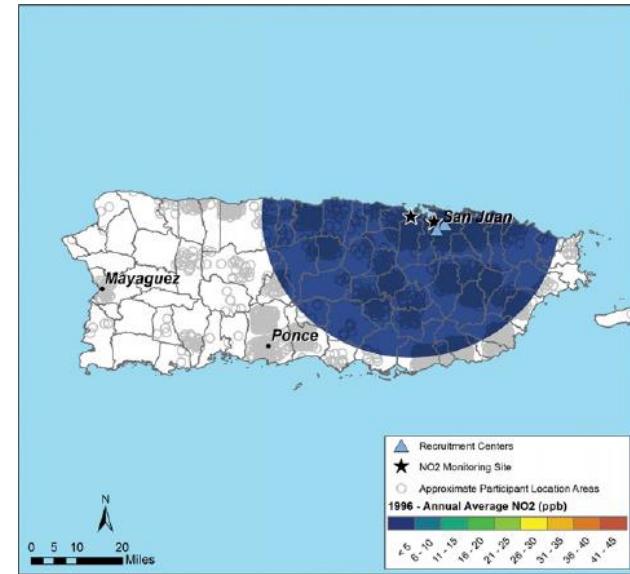
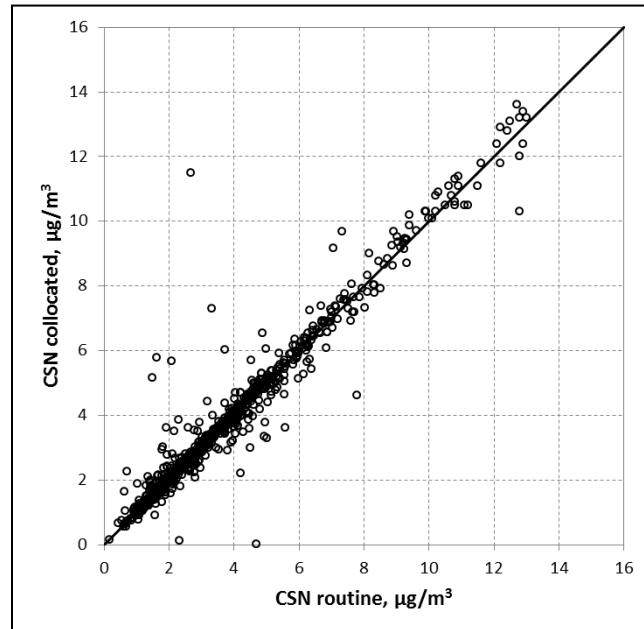
Intraurban Variability of PM_{2.5} Components across Metropolitan Saint Louis

**Li Du, Kelsey Haddad, Jay Turner
Washington University in St. Louis**

Aug 2014

Intraurban Variability of PM_{2.5} Components in the Context of Measurement Error

- One motivation: epidemiological studies using one central site to represent the entire area of interest may lead to exposure misclassification



Nishimura, K.K., et al., 2013. *American Journal of Respiratory and Critical Care Medicine* 188, 309-318

- Measurement error may contribute to temporal and spatial variability estimates

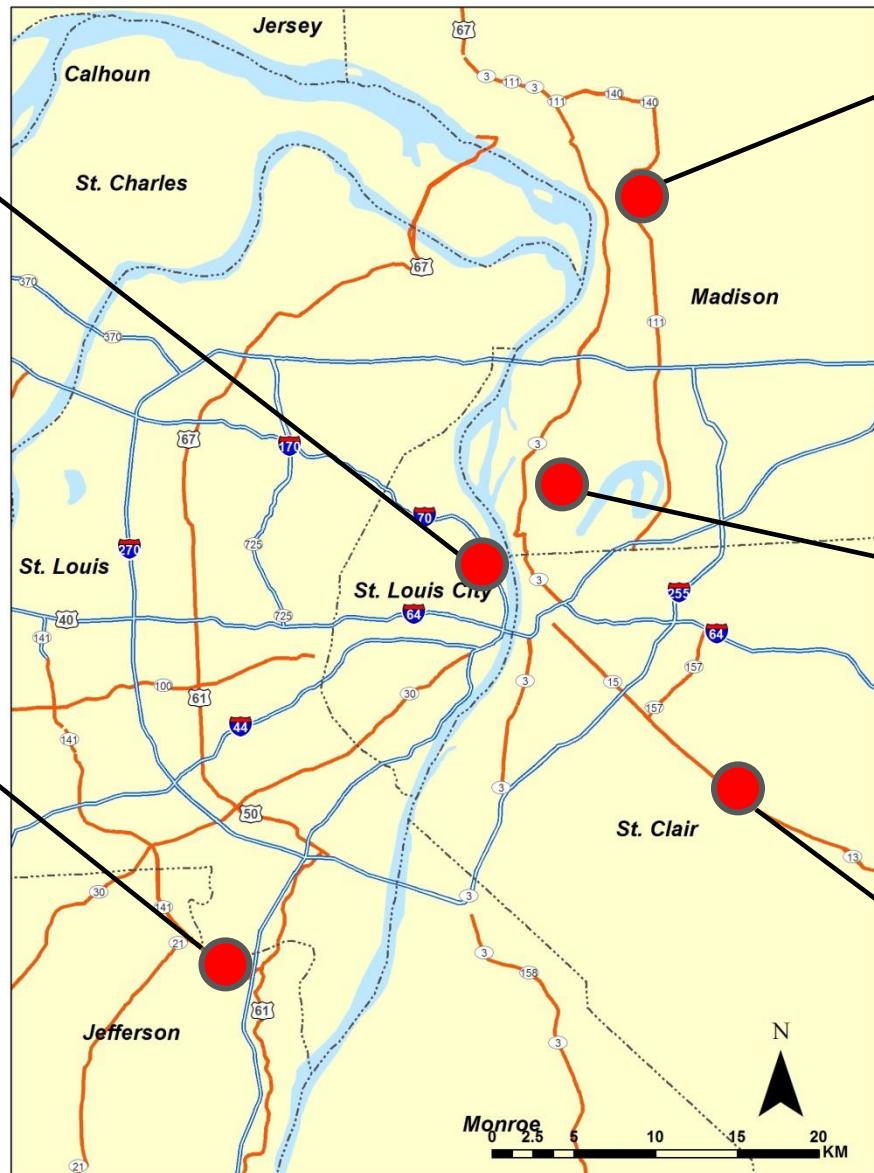
Where We Currently Measure PM_{2.5} Components in St. Louis

Blair Street
(City of St. Louis)

STN site, NATTS site
Start Feb 2000
MDNR

Arnold West

CSN site
Start Apr 2000
MDNR



Roxana
CSN protocol site
Start Jul 2012
Wash Univ. STL

Granite City
CSN site
Start Oct 2007

Belleville
CSN protocol site
Start Feb 2010
ARA, Inc.

Common Metrics To Gauge Spatial Variability

- Pearson correlation coefficient (r)
- Coefficient of divergence (COD)
- Scattergram and linear regression

Pearson Correlation Coefficient (r)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

X_i = site 1 conc., sample i

Y_i = site 2 conc., sample i

- Describes the temporal variability between two sites
- Range (-1, +1)
 - -1 : perfect negative correlation
 - +1: perfect positive correlation
 - 0 : no correlation
- Very sensitive to outliers/extreme values

Species Correlation Coefficients Across the Five Sites

	BL - AR	BL - GC	BL - BV	BL - RQ	AR - GC	AR - BV	AR - RQ	BV - RQ	BV - GC
PM _{2.5} mass	0.84	0.74	0.80	0.90	0.90	0.80	0.83	0.88	0.66
SO ₄	0.96	0.96	0.96	0.90	0.90	0.93	0.82	0.90	0.91
NO ₃	0.97	0.96	0.97	0.92	0.93	0.97	0.93	0.96	0.93
NH ₄	0.96	0.94	0.96	0.94	0.89	0.94	0.91	0.94	0.91
OC_TOR	0.88	0.69	0.82	0.88	0.80	0.83	0.91	0.89	0.78
EC_TOR	0.77	0.77	0.56	0.54	0.58	0.66	0.68	0.74	0.41
Si	0.71	0.46	0.90	0.78	0.52	0.91	0.84	0.88	0.56
Fe	0.20	-0.13	0.21	0.05	0.13	0.21	0.42	0.31	0.02
Ca	0.17	0.30	0.66	0.42	0.33	0.58	0.34	0.60	0.49

AR – Arnold

BL – Blair Street (City of St. Louis)

BV – Belleville

GC – Granite City

RQ – Roxana

Includes all data above detection limits

Species Correlation Coefficients Across the Five Sites

	BL - AR	BL - GC	BL - BV	BL - RQ	AR - GC	AR - BV	AR - RQ	BV - RQ	BV - GC
PM _{2.5} mass	0.84	0.74	0.80	0.90	0.90	0.80	0.83	0.88	0.66
SO ₄	0.96	0.96	0.96	0.90	0.90	0.93	0.82	0.90	0.91
NO ₃	0.97	0.96	0.97	0.92	0.93	0.97	0.93	0.96	0.93
NH ₄	0.96	0.94	0.96	0.94	0.89	0.94	0.91	0.94	0.91
OC_TOR	0.88	0.69	0.82	0.88	0.80	0.83	0.91	0.89	0.78
EC_TOR	0.77	0.77	0.56	0.54	0.58	0.66	0.68	0.74	0.41
Si	0.71	0.46	0.90	0.78	0.52	0.91	0.84	0.88	0.56
Fe	0.20	-0.13	0.21	0.05	0.13	0.21	0.42	0.31	0.02
Ca	0.17	0.30	0.66	0.42	0.33	0.58	0.34	0.60	0.49

AR – Arnold

BL – Blair Street (City of St. Louis)

BV – Belleville

GC – Granite City

RQ – Roxana

Includes all data above detection limits

Species Correlation Coefficients Across the Five Sites

	BL - AR	BL - GC	BL - BV	BL - RQ	AR - GC	AR - BV	AR - RQ	BV - RQ	BV - GC
PM _{2.5} mass	0.84	0.74	0.80	0.90	0.90	0.80	0.83	0.88	0.66
SO ₄	0.96	0.96	0.96	0.90	0.90	0.93	0.82	0.90	0.91
NO ₃	0.97	0.96	0.97	0.92	0.93	0.97	0.93	0.96	0.93
NH ₄	0.96	0.94	0.96	0.94	0.89	0.94	0.91	0.94	0.91
OC_TOR	0.88	0.69	0.82	0.88	0.80	0.83	0.91	0.89	0.78
EC_TOR	0.77	0.77	0.56	0.54	0.58	0.66	0.68	0.74	0.41
Si	0.71	0.46	0.90	0.78	0.52	0.91	0.84	0.88	0.56
Fe	0.20	-0.13	0.21	0.05	0.13	0.21	0.42	0.31	0.02
Ca	0.17	0.30	0.66	0.42	0.33	0.58	0.34	0.60	0.49

AR – Arnold

BL – Blair Street (City of St. Louis)

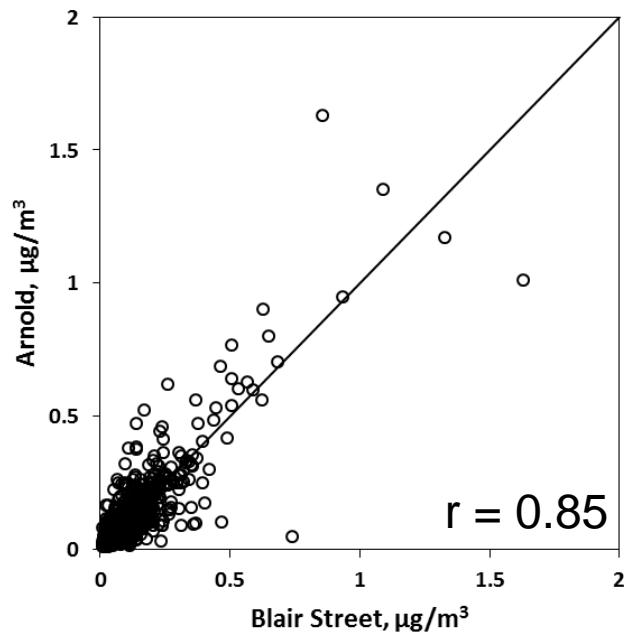
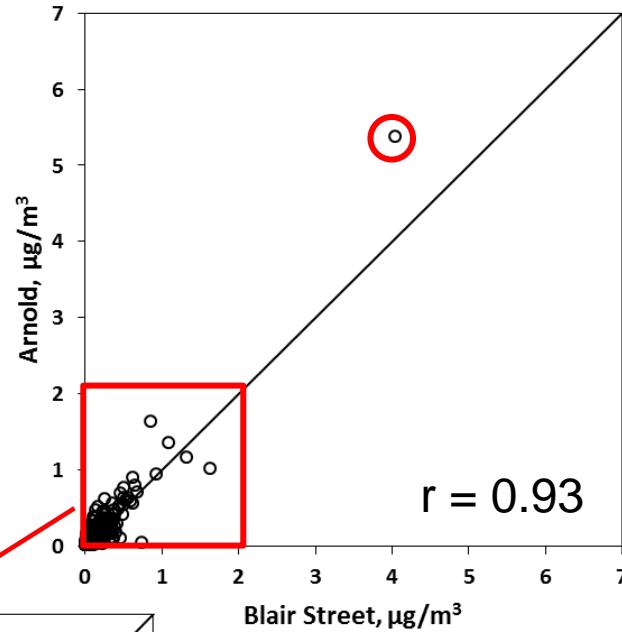
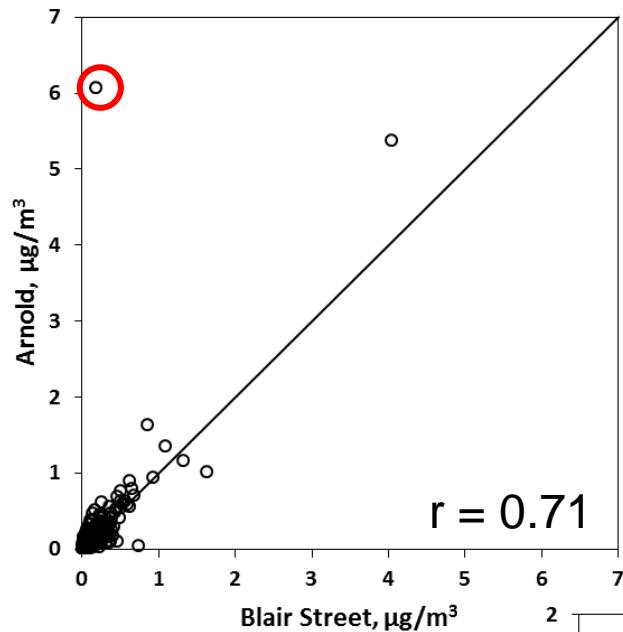
BV – Belleville

GC – Granite City

RQ – Roxana

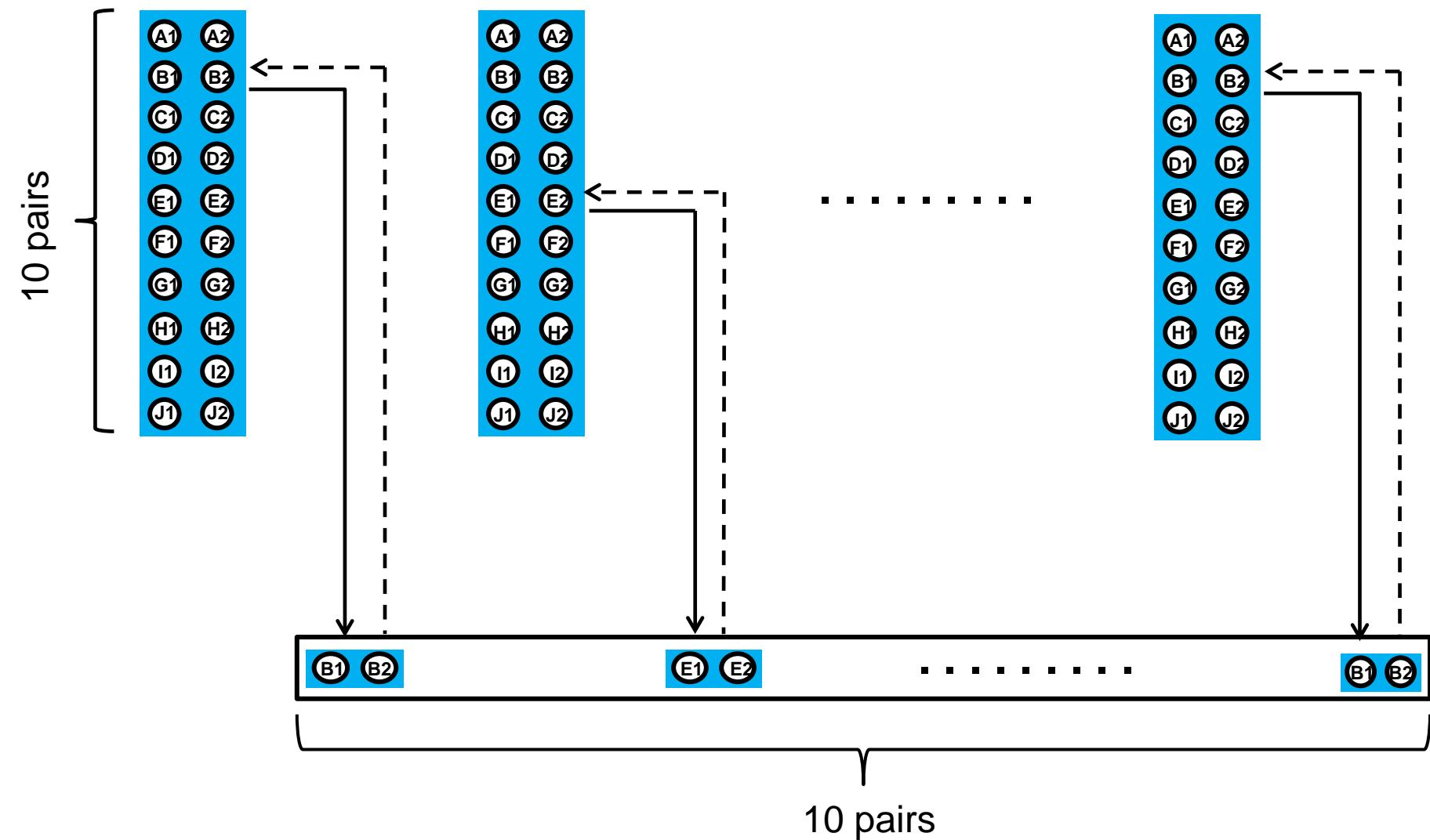
Includes all data above detection limits

Sensitivity to Outliers/Extreme Values - Silicon

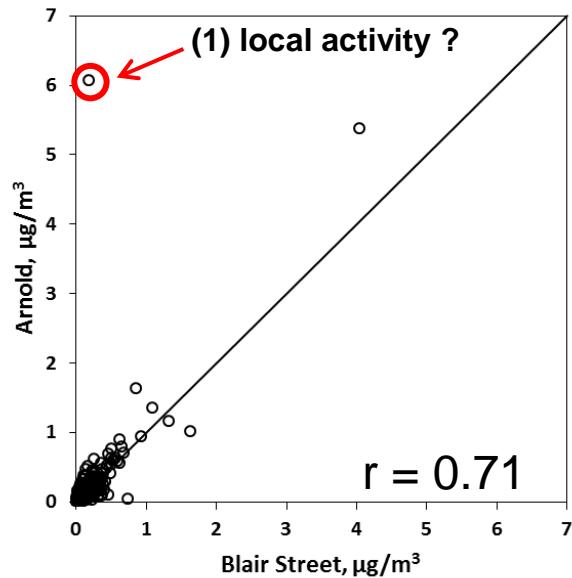


Bootstrapping

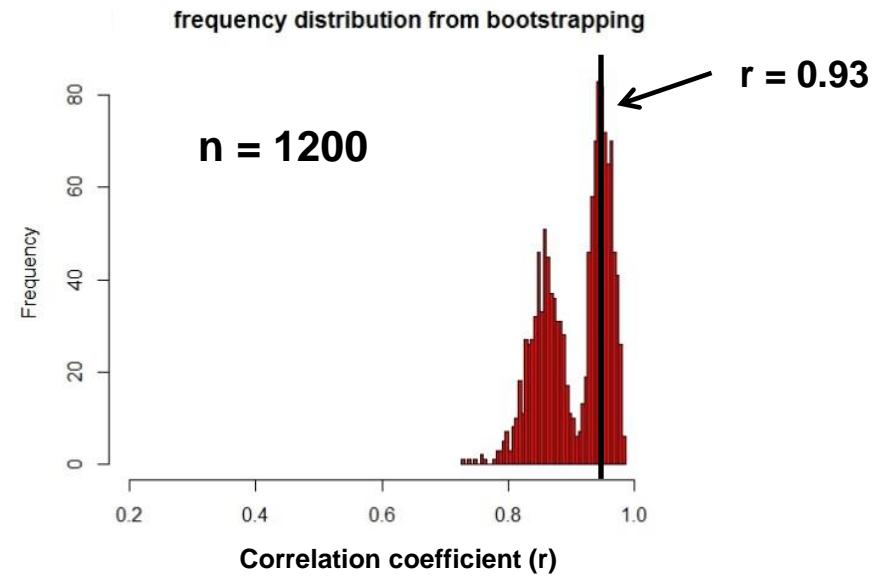
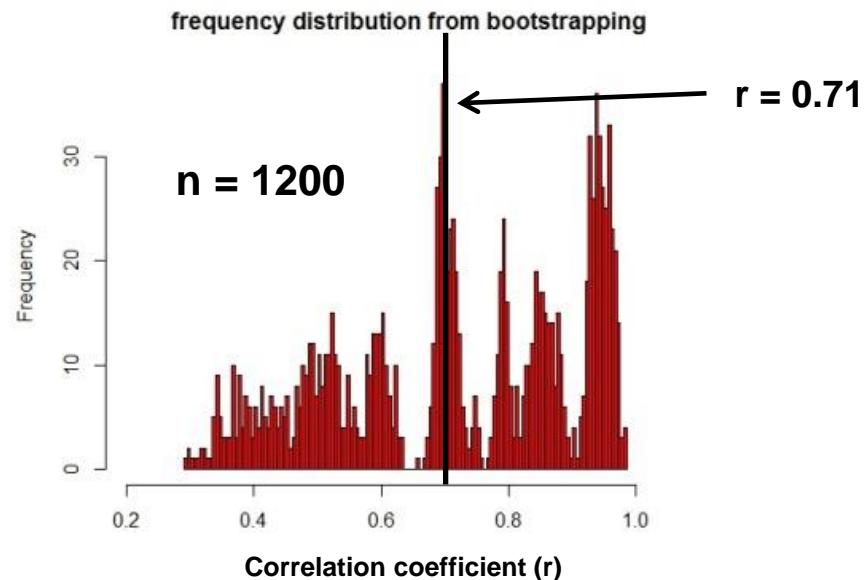
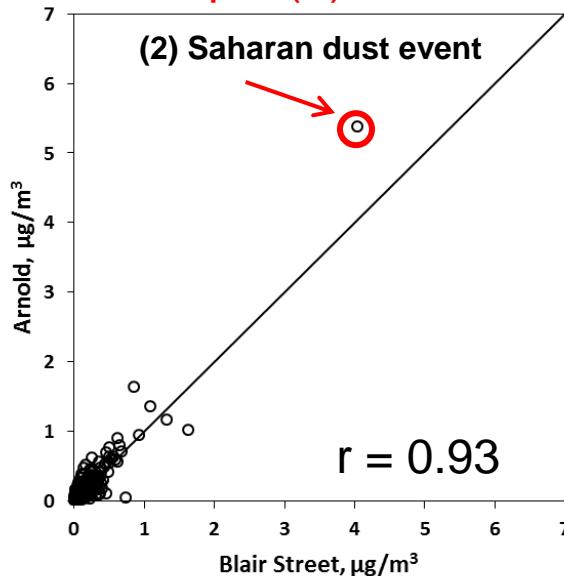
Resampling technique: data pairs are drawn randomly from the original dataset **WITH REPLACEMENT** to create a new dataset.



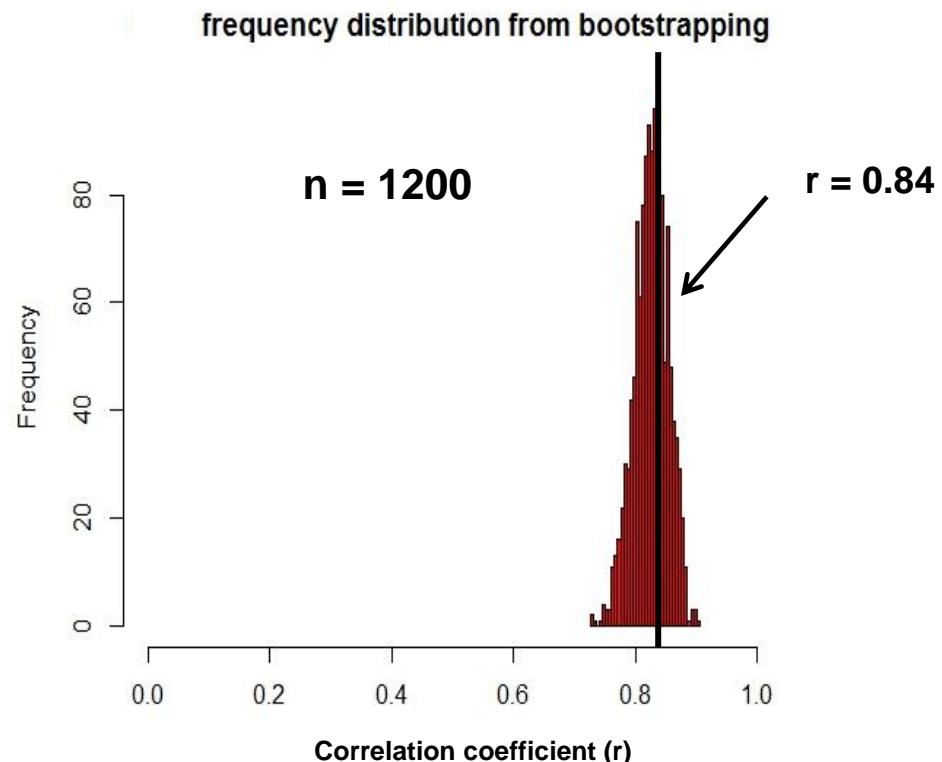
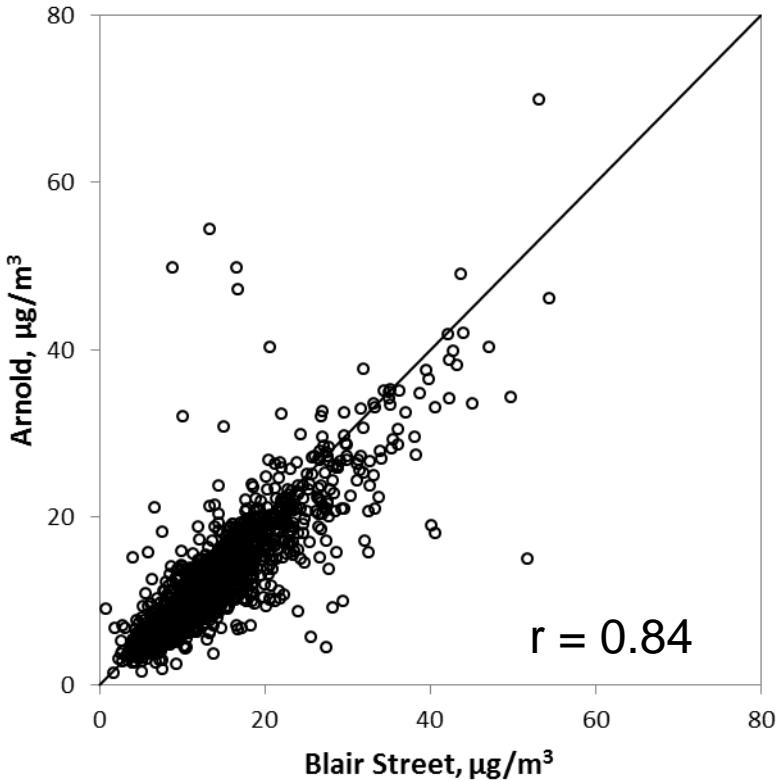
PM_{2.5} Silicon: Blair Street vs. Arnold



Remove sample (1)...



PM_{2.5} Mass: Blair Street vs. Arnold



Caution must be used when applying the correlation coefficient because a few extreme values may significantly influence the result.

Coefficient of Divergence (COD)

$$\text{COD} = \sqrt{\frac{1}{p} \sum_{i=1}^p \left(\frac{X_i - Y_i}{X_i + Y_i} \right)^2}$$

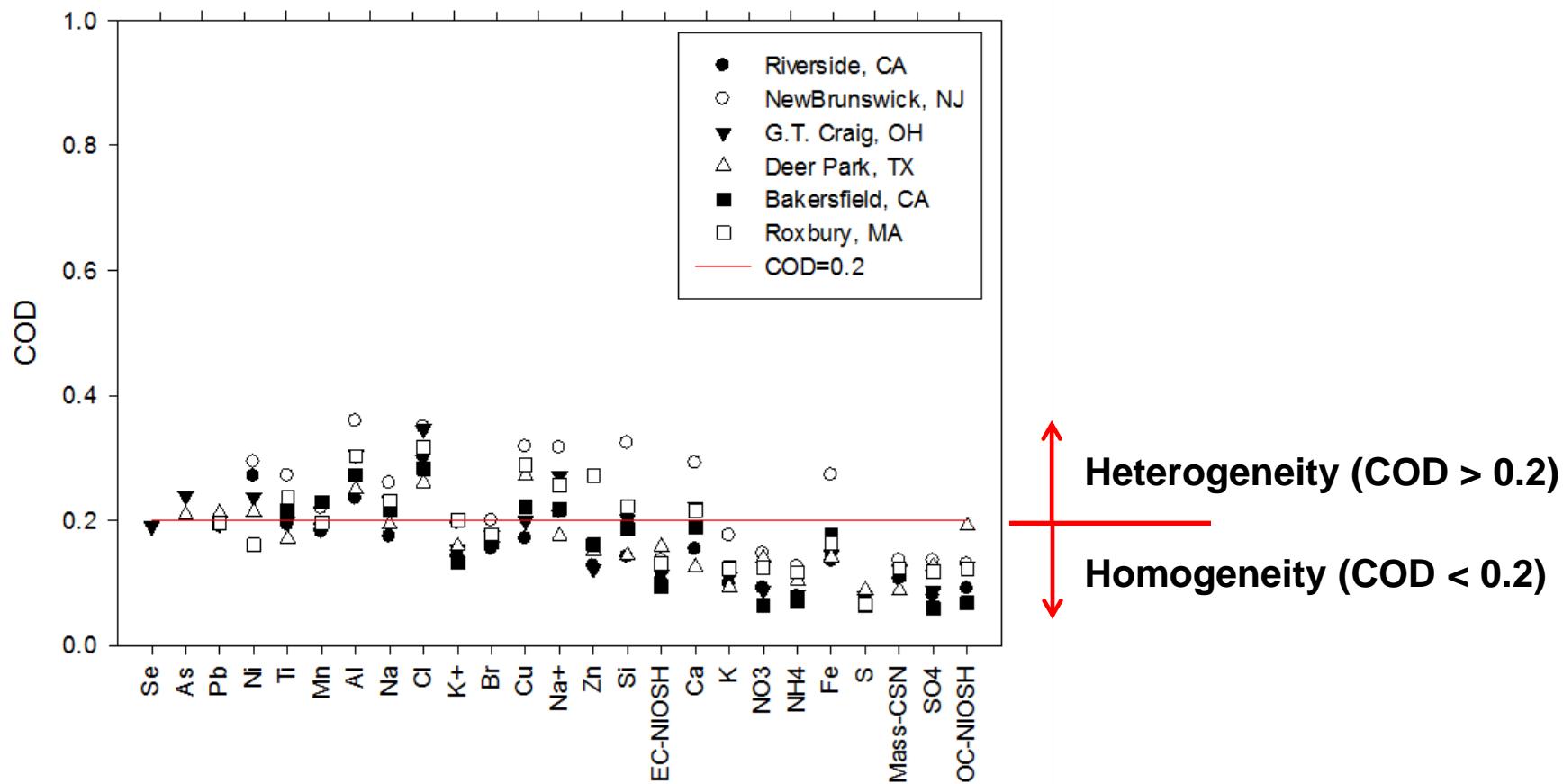
X_i = site 1 conc., sample i

Y_i = site 2 conc., sample i

- Captures the spatial variability between two sites
- Range (0, 1)
 - 0: perfect homogeneity
 - 1: perfect heterogeneity
- Less sensitive to outliers/extreme values but...

Spatial Variability in the Context of Measurement Error

COD for PM_{2.5} species at the Six CSN Collocation Sites



Species Coefficient of Divergence Across the Five Sites

	Collo Sites	BL - AR	BL - GC	BL - BV	BL - RQ	AR - GC	AR - BV	AR - RQ	BV - RQ	BV - GC
PM _{2.5} mass	0.09 - 0.14	0.15	0.16	0.16	0.13	0.15	0.20	0.17	0.16	0.19
SO ₄	0.06 - 0.14	0.10	0.12	0.12	0.16	0.15	0.12	0.21	0.16	0.18
NO ₃	0.06 - 0.15	0.20	0.15	0.15	0.19	0.24	0.12	0.23	0.17	0.17
NH ₄	0.07 - 0.13	0.18	0.14	0.15	0.20	0.18	0.21	0.28	0.19	0.16
OC_TOR	0.06 - 0.17	0.13	0.15	0.16	0.12	0.18	0.16	0.12	0.13	0.17
EC_TOR	0.07 - 0.19	0.29	0.22	0.27	0.24	0.30	0.29	0.31	0.21	0.27
Si	0.14 - 0.32	0.22	0.33	0.20	0.24	0.30	0.16	0.21	0.19	0.34
Fe	0.13 - 0.27	0.46	0.63	0.41	0.43	0.66	0.26	0.29	0.30	0.73
Ca	0.12 - 0.29	0.25	0.39	0.27	0.34	0.40	0.27	0.36	0.29	0.47

AR – Arnold

BL – Blair Street (City of St. Louis)

BV – Belleville

GC – Granite City

RQ – Roxana

Includes all data above detection limits

Species Coefficient of Divergence Across the Five Sites

	Collo Sites	BL - AR	BL - GC	BL - BV	BL - RQ	AR - GC	AR - BV	AR - RQ	BV - RQ	BV - GC
PM _{2.5} mass	0.09 - 0.14	0.15	0.16	0.16	0.13	0.15	0.20	0.17	0.16	0.19
SO ₄	0.06 - 0.14	0.10	0.12	0.12	0.16	0.15	0.12	0.21	0.16	0.18
NO ₃	0.06 - 0.15	0.20	0.15	0.15	0.19	0.24	0.12	0.23	0.17	0.17
NH ₄	0.07 - 0.13	0.18	0.14	0.15	0.20	0.18	0.21	0.28	0.19	0.16
OC_TOR	0.06 - 0.17	0.13	0.15	0.16	0.12	0.18	0.16	0.12	0.13	0.17
EC_TOR	0.07 - 0.19	0.29	0.22	0.27	0.24	0.30	0.29	0.31	0.21	0.27
Si	0.14 - 0.32	0.22	0.33	0.20	0.24	0.30	0.16	0.21	0.19	0.34
Fe	0.13 - 0.27	0.46	0.63	0.41	0.43	0.66	0.26	0.29	0.30	0.73
Ca	0.12 - 0.29	0.25	0.39	0.27	0.34	0.40	0.27	0.36	0.29	0.47

AR – Arnold

BL – Blair Street (City of St. Louis)

BV – Belleville

GC – Granite City

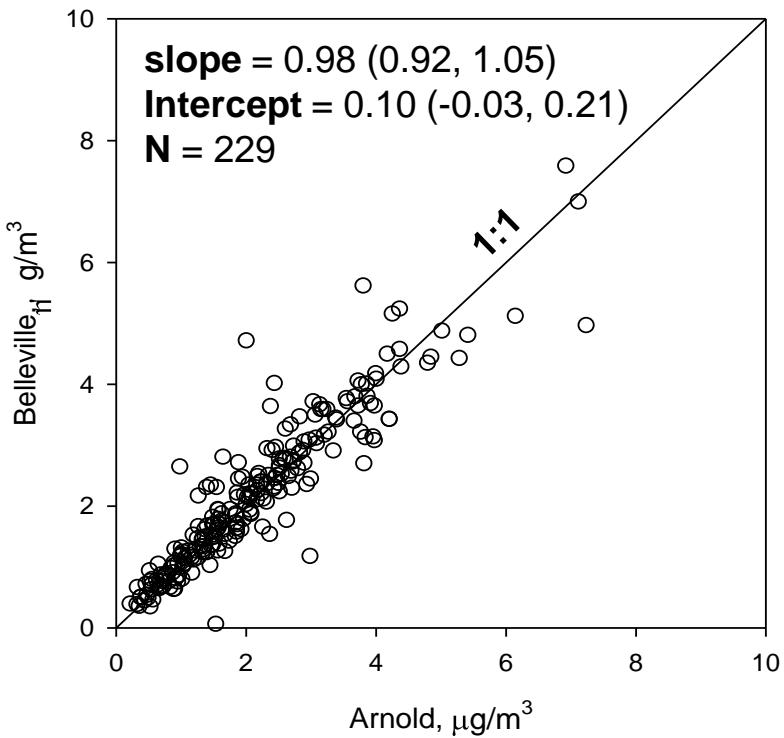
RQ – Roxana

Includes all data above detection limits

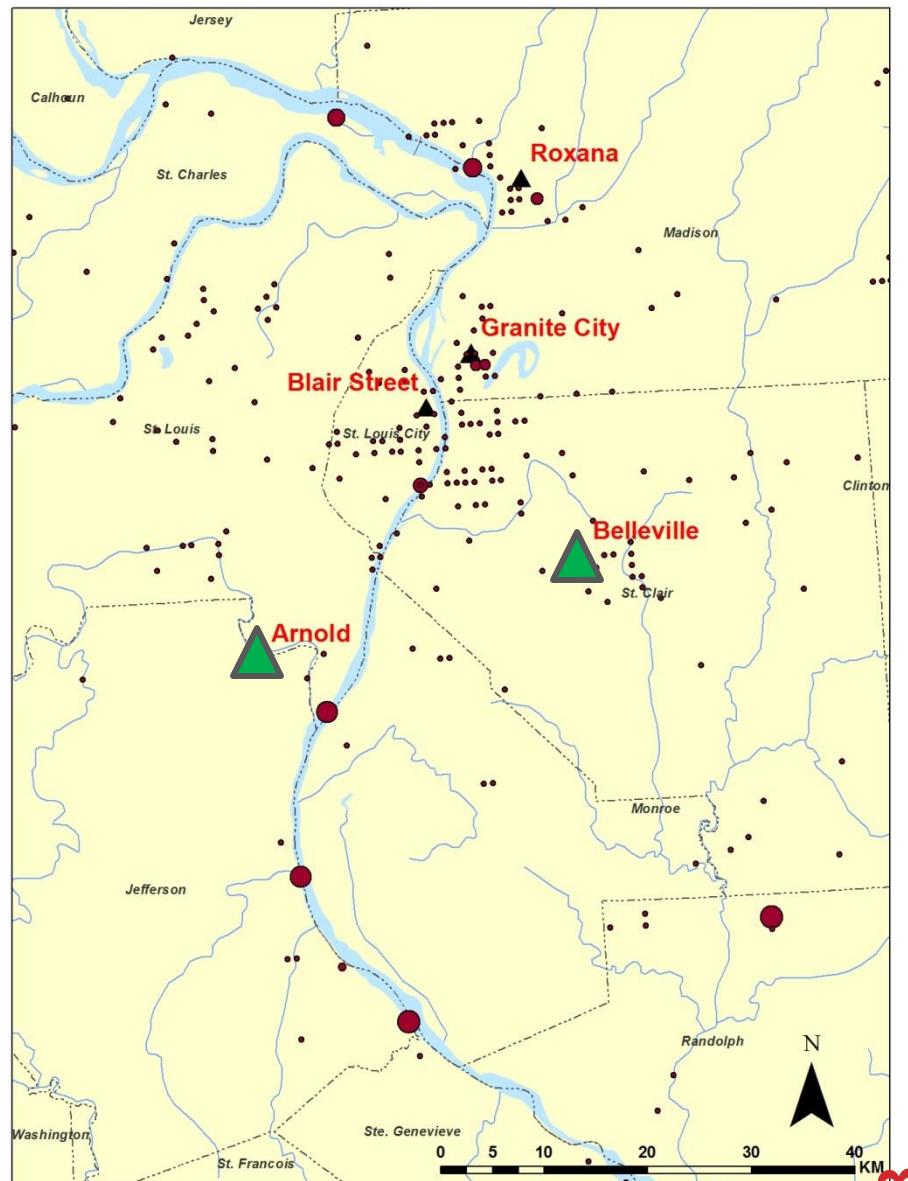
Scattergram/Linear Regression

- **Simple graphical method**
- **Slope of unity and intercept of zero often used to infer spatial and temporal homogeneity**

PM_{2.5} Sulfate: Belleville vs. Arnold



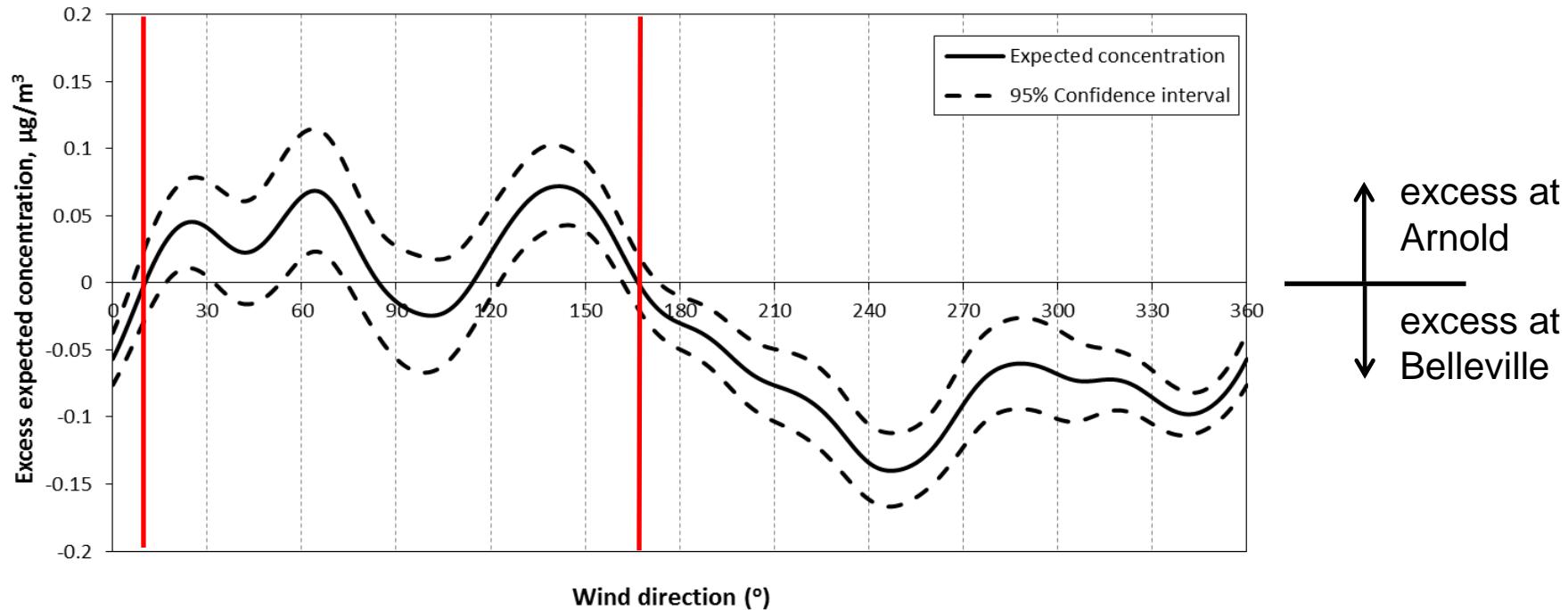
- Linear regression and COD = **0.12** (< 0.2) suggest spatial homogeneity of sulfate across the two sites
- However, there is more scatter than is predicted from collocated precision at the six CSN collocation sites
 - Local sources differentially impacting the sites?



circles: 2011 NEI SO₂ point sources

Non-parametric Wind Regression (NWR): Arnold – Belleville Sulfate Conc. Difference

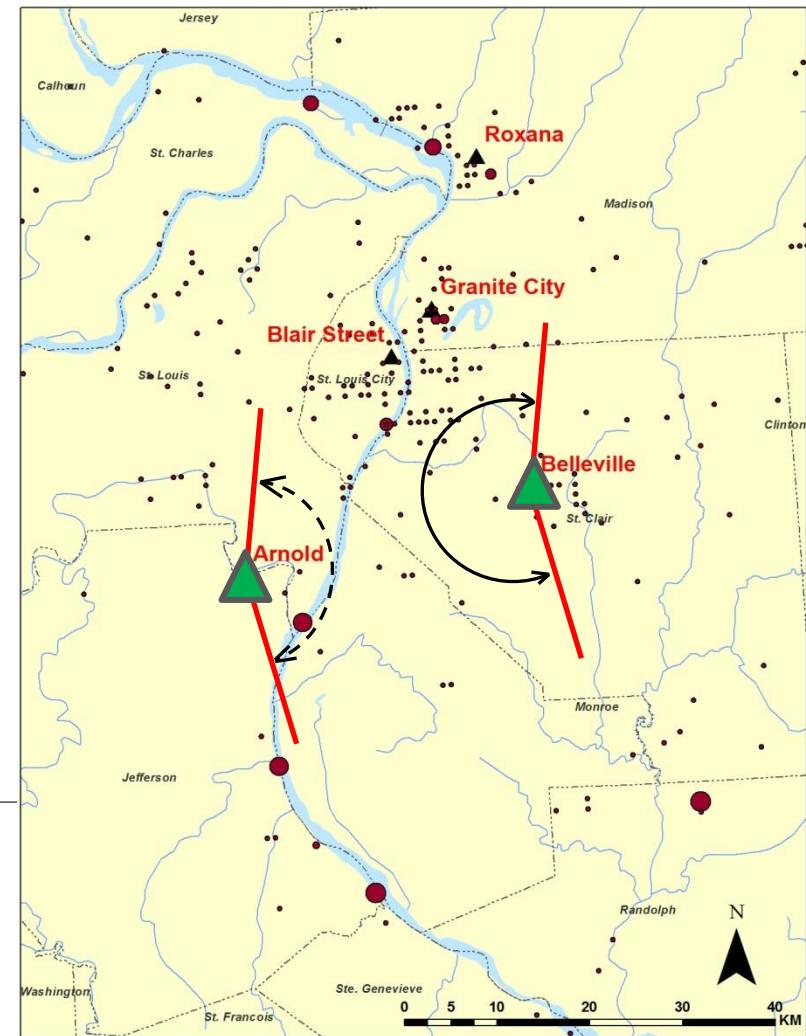
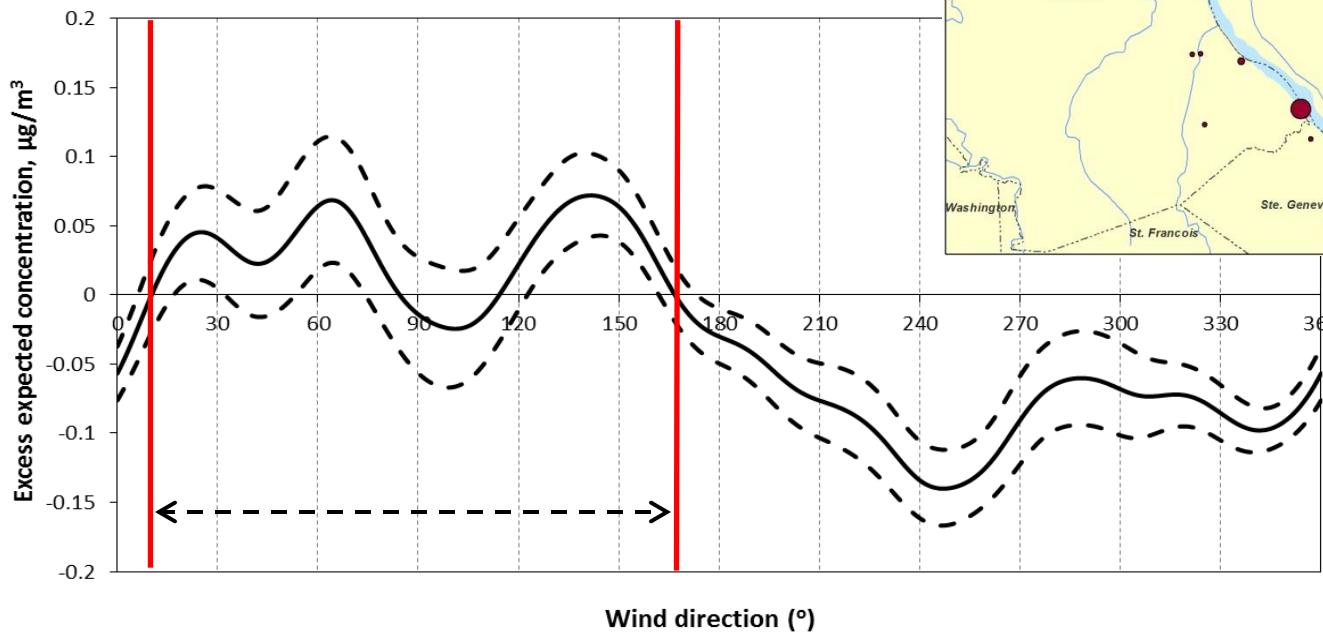
NWR is essentially a pollution rose calculation but with no *a priori* binning of wind directions and with better statistical support



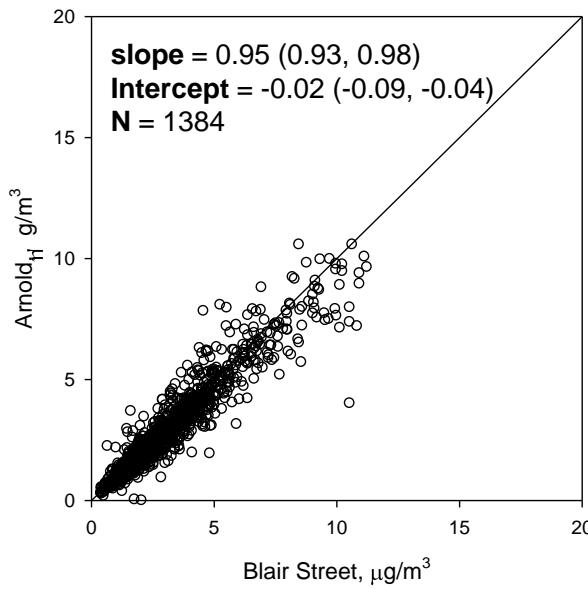
Mean differences are relatively small but have a consistent pattern

Sulfate Conc. Difference between Two Sites

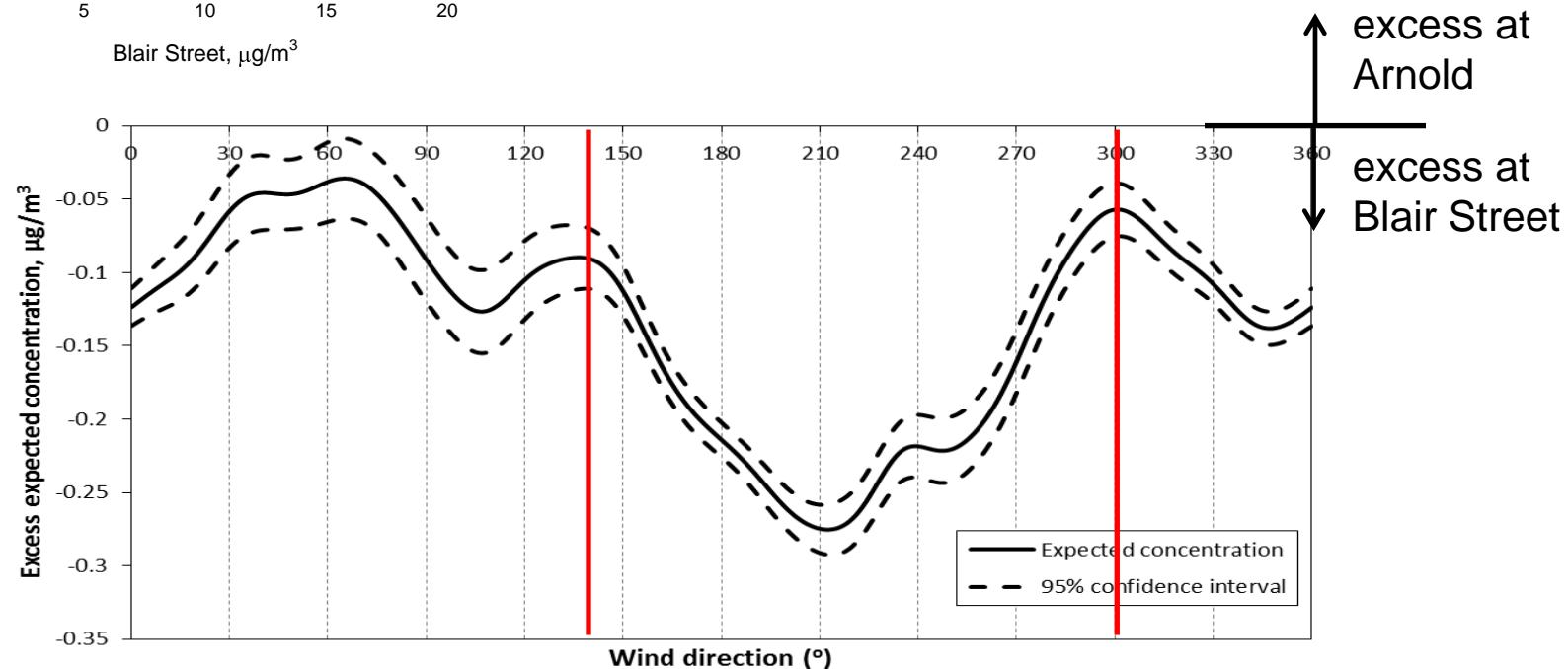
- Sulfate higher at the site impacted by urban core and industrial zones along the riverfront



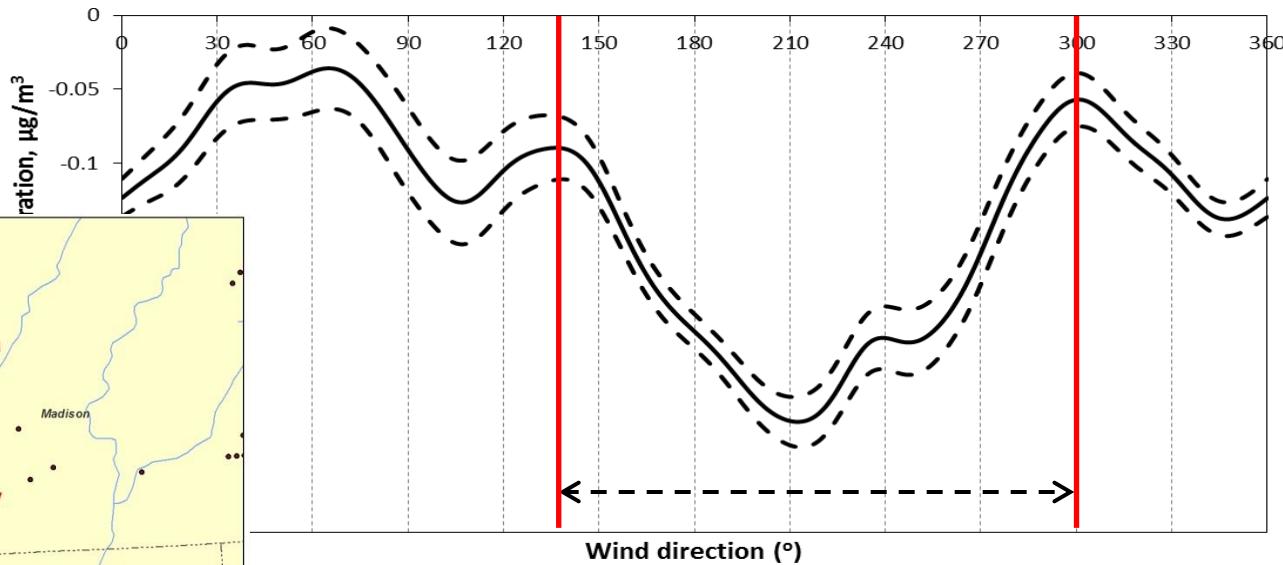
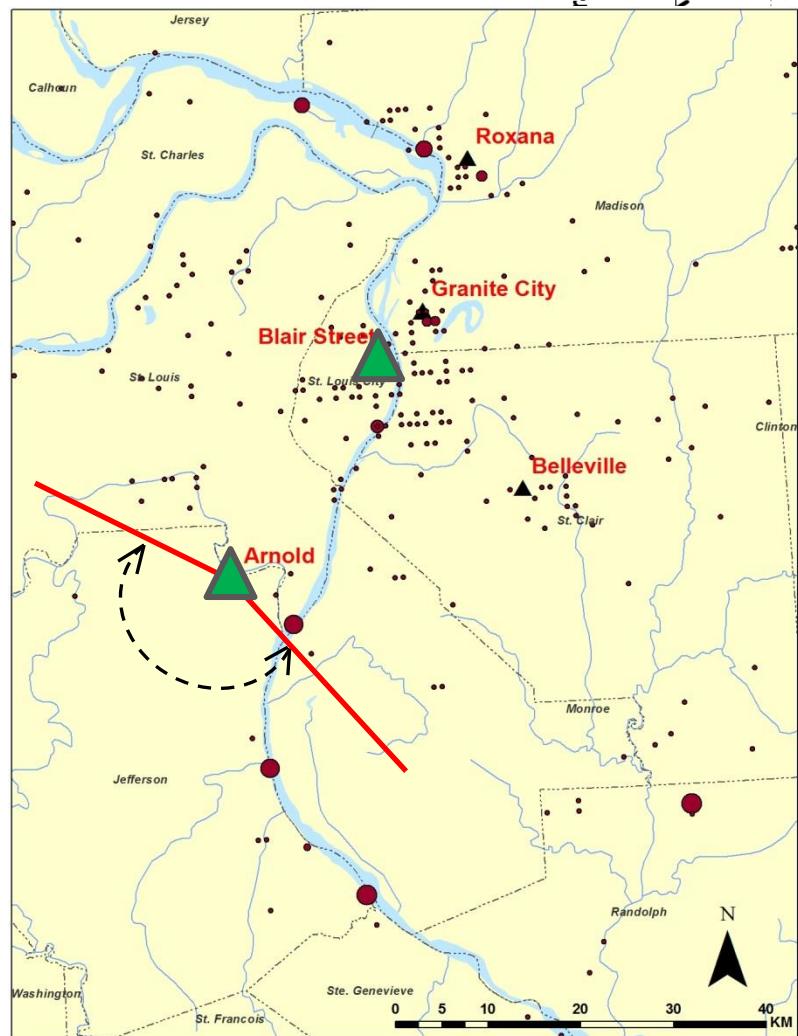
Non-parametric Wind Regression (NWR): Arnold – Blair Street Sulfate Conc. Difference



- Sulfate higher at Blair (City of St. Louis), mean differences is $0.14 \mu\text{g}/\text{m}^3$
- Excess sulfate at Blair for all wind directions but there is a pattern



Sulfate Conc. Difference between Two Sites



- Maximum excess sulfate at Blair when Arnold is upwind of the urban core and industrial zones along the riverfront

Summary & Conclusions

- **Use caution when characterizing spatial and temporal variability by simple metrics**
- **Carefully examine and condition the dataset before the metrics are applied**
- **Consider the contribution of measurement errors to observed spatial and temporal variability**
 - **What portion of the observed variation can it explain?**
- **Sometimes sources (or source zones) with even relatively small impacts can be teased out of the intersite variability**

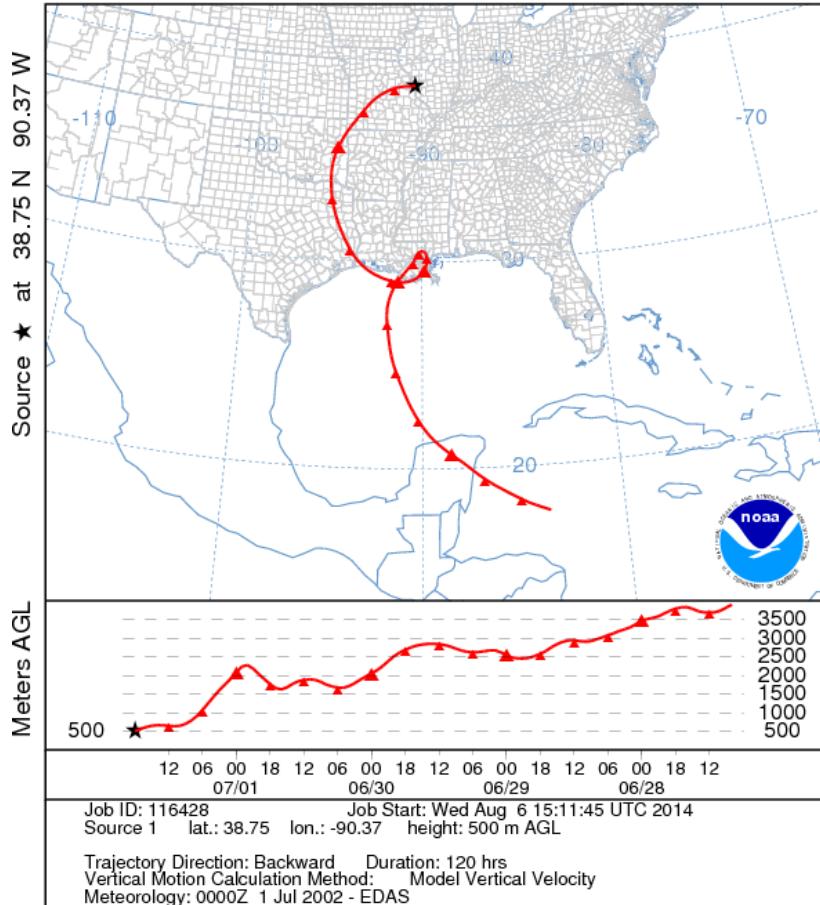
Acknowledgements

Data Sources

- Missouri DNR
- Illinois EPA
- ARA, Inc. (Eric Edgerton)

Roxana measurements are funded by Phillips66

NOAA HYSPLIT MODEL
Backward trajectory ending at 1800 UTC 01 Jul 02
EDAS Meteorological Data



NOAA HYSPLIT MODEL
Backward trajectory ending at 1800 UTC 30 Jun 05
EDAS Meteorological Data

